DOCUMENT RESUME

ED 236 194 TM 830 703

AUTHOR Tindal, Gerald; Deno, Stanley
TITLE Factors Influencing the Agreement between Visual and Statistical Analyses of Time Series Data.
INSTITUTION Minnesota Univ., Minneapolis. Inst. for Research on Learning Disabilities.
SPONS AGENCY Office of Special Education and Rehabilitative Services (ED), Washington, DC.
REPORT NO IRLD-RR-118
PUB DATE Jun 83
CONTRACT 300-80-0622
NOTE 40p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Adults; *Correlation; *Data Analysis; Elementary School Teachers; Elementary Secondary Education; Graphs; Higher Education; Inservice Teacher Education; Preservice Teacher Education; Research Methodology; Secondary School Teachers; *Statistical Analysis; *Validity; Visual Discrimination; *Visual Perception
IDENTIFIERS *Data Interpretation; *Time Series Analysis

ABSTRACT

The focus of this study was on the relationship between visual and statistical analyses of time series data and the degree to which characteristics of the data influenced this relationship. A total of 52 teachers took part in evaluating a series of graphs having prespecified characteristics. The independent variables manipulated included: slope, variability, training, and aimline/decision rules. Generally, the influence of these variables was significant and in the predicted direction. However, the overall level of relationship between the two analytic procedures was modest. The implications of this research are discussed in terms of the manner in which visual analysis is conducted and the procedures needed to establish statistical conclusion validity. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED236194

University of Minnesota

Research Report No. 118

FACTORS INFLUENCING THE AGREEMENT BETWEEN VISUAL AND STATISTICAL ANALYSES OF TIME SERIES DATA

Gerald Tindal and Stanley Deno

SCOPE OF INTEREST NOTICE
The ERIC Facility has assigned this document for processing to: TM EC
In our judgement, this document is also of interest to the clearinghouses noted to the right. Indexing should reflect their special points of view.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
J. Ysseldyke
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Institute for Research on Learning Disabilities

TM 830 703

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

IRLD

Director: James E. Ysseldyke

The Institute for Research on Learning Disabilities is supported by a contract (300-80-0622) with the Office of Special Education, Department of Education, through Title VI-G of Public Law 91-230. Institute investigators are conducting research on the assessment/decision-making/intervention process as it relates to learning disabled students.

During 1980-1983, Institute research focuses on four major areas:

- Referral
- Identification/Classification
- Intervention Planning and Progress Evaluation
- Outcome Evaluation

Additional information on the Institute's research objectives and activities may be obtained by writing to the Editor at the Institute (see Publications list for address).

The research reported herein was conducted under government sponsorship. Contractors are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent the official position of the Office of Special Education.

Research Report No. 118

FACTORS INFLUENCING THE AGREEMENT BETWEEN VISUAL AND STATISTICAL ANALYSES OF TIME SERIES DATA

Gerald Tindal and Stanley Deno

Institute for Research on Learning Disabilities

University of Minnesota

June, 1983

Abstract

The focus of this study was on the relationship between visual and statistical analyses of time series data and the degree to which characteristics of the data influenced this relationship. A total of 52 subjects took part in evaluating a series of graphs having pre-specified characteristics. The independent variables manipulated included: slope, variability, training, and aimline/decision rules. Generally, the influence of these variables was significant and in the predicted direction. However, the overall level of relationship between the two analytic procedures was modest. The implications of this research are discussed in terms of the manner in which visual analysis is conducted and the procedures needed to establish statistical conclusion validity.

Factors Influencing the Agreement Between Visual and Statistical Analyses of Time Series Data

For nearly 10 years now, there has been a controversy in the behavioral literature over the appropriate analysis of time series data. Many arguments have been presented both for and against the use of statistics in such analyses. However, relatively few studies have been conducted to investigate the use of statistical analysis. In part, this has been due to the unique characteristics of time series data, which often make it difficult to use any procedures other than visual analysis.

While classical statistical procedures developed from R. A. Fisher (1951) have been discounted (cf. Journal of Applied Behavior Analysis, v. 7, No. 4, 1974), several alternative procedures have been proposed, including statistical analyses utilizing time series analysis (Glass, Willson, & Gottman, 1975; Gottman, 1973; Gottman & Glass, 1978), randomization tests (Edgington, 1967, 1972a, 1972b), the Rn statistic (Revusky, 1967), and the c-statistic (Tryon, 1982). The use of statistics in analyzing single subject data must address several issues that are not relevant in the more traditional between-group analysis. A central problem in the case of repeated measurement over time is the degree to which successive data points are related to each other. This characteristic usually is referred to as serial dependency or autocorrelation in which there is "a correlation (r) between data points separated by different time intervals (lags) in the series" (Kazdin, 1976, p. 273).

In statistics using an analysis of variance model (F, t), the assumption of independence of error components is critical and if

violated, precludes the use of such techniques as an appropriate alternative. In time series data, there is considerable serial dependency present and as a result the error components cannot be assumed to be independent. In an empirical analysis, Kratochwill, Alden, Demuth, Dawson, Panicucci, Arntson, McMurray, Hempstead, and Levin (1974) demonstrated that the assumption of statistical independence necessary for use of ANOVA is unwarranted in an N=1 design. For example, given two behavioral outcomes (e.g., on-task and off-task), the probability of any one occurrence is dependent upon previous occurrences, both in terms of the number of times the sequence changes from one outcome to another (from on-task to off-task) as well as the number of consecutive observations characterized by the same behavior (the length of strings of on-task and off-task occurrences).

The presence of serial dependency has other effects on conventional analyses. First, the number of independent sources of information in the data is reduced, resulting in an overestimate of the true value of F or t. Second, there is a spurious reduction in the variability of the data, resulting in an underestimate of the variability that would have been obtained from independent observations, the net effect of which is a positively biased F or t (Kazdin, 1976). Third, Type I error is underestimated for positive autocorrelation and overestimated for negative autocorrelation (Scheffe, 1959). The use of ANOVA procedures on time series data will also result in the magnitude of the mean square (MS) error being greatly increased and, as a consequence, the probability of detecting

a true treatment effect will be decreased. Error variance in this model is calculated from the deviation of scores within conditions from the condition mean, with no account taken of trend in the data. When all data points are included in determining the mean, rather than using only those obtained at the point at which asymptotic levels of performance have been reached (the peak level of responding), the problem cited above, increasing the magnitude of the MS error, is exacerbated (Hartmann, 1974).

Gottman and Glass (1978), in arguing for the use of statistics, countered that the large effects that operant psychologists are accustomed to detecting are the unique phenomenon of the experimental laboratory. In applied settings, such as schools and hospitals, there is little control over many of the environmental stimuli, and as a consequence, "one has every reason to expect small effects outside the laboratory" (p. 199). Therefore, they argued that small effects detectable by sensitive statistical procedures, but undetectable by visual analysis, are important to know about if further research and experimental manipulations are to be investigated. Kazdin (1975) agreed, noting that interventions that show only a modest effect alone may be important when added to other interventions. Other situations in which some argue for the use of statistical procedures include: (a) instances where it is difficult to achieve a stable baseline and time or ethical constraints preclude further waiting; (b) when visual analysis is equivocal and the effects of changes are ambiguous; and (c) in applied settings characterized by uncontrolled variation (Kazdin, 1976).

According to Glass et al. (1975), data from time series experiments that do not appear statistically significant when visually inspected, often turn out to be significant when "appropriately tested" (p. 62). When Jones, Weinrott, and Vaught (1975) reanalyzed the Hall, Fox, Willard, Goldsmith, Emerson, Owen, Davis, and Porcia (1971) study, they found that different conclusions might be reached in relying on visual and statistical criteria for evaluating change in trend. Here, changes in slope "appeared" to be significant but were not when analyzed using statistical techniques.

Gottman and Glass (1978) investigated the degree to which visual analysis and statistical analysis were in agreement. Thirteen graduate students in a seminar on time-series analysis were asked to inspect various graphs and judge whether or not an intervention effect was present. The same data were analyzed by methods outlined in detail by Glass et al. (1975), as a first order integrated moving averages process, and a t statistic applied for each intervention. In sum, they found the "eyeball test" to give results that varied from judge to judge and were in sharp conflict with the findings of statistical tests.

Jones, Vaught, and Weinrott (1977) reanalyzed several investigations reported in the Journal of Applied Behavior Analysis (JABA) representing a great variety of scores and design properties, using time-series analysis. In some instances, time-series analysis corroborated the author's visually based conclusions; in others the two analyses yielded different conclusions; and in still others, the time-series analysis revealed findings not apparent in the authors' visual analysis.

Finally, an investigation was conducted by Jones, Weinrott, and Vaught (1978) to ascertain to extent to which serial dependency influenced the agreement between inferences based on visual or time-series analysis. JABA graphs were presented to judges well versed in behavior charting and they were asked whether or not a meaningful change in level had been demonstrated from one phase to another. Graphs were selected by the authors in which the effects were sufficiently "nonobvious" to warrant critical analysis, and serial dependency was apparent. The graphs were further blocked into three different levels of serial dependency by two levels of significance of difference in level between phases.

Their results indicated that agreement between visual analysis and time-series analysis was inversely related to the magnitude of the serial dependency in the scores. That is, the more serial dependency present, the less reliable visual analysis tended to be. Furthermore, they found that visual and time-series inferences agreed better when no statistical changes in level were present. Finally, an interaction effect was present in which visual and time-series inferences were most in agreement when the data showed neither serial dependency nor significant differences in level. In effect, judges tended to agree with time-series analysis that no effect was present but disagreed most when an effect was present. Intercorrelations among the 11 judges ranged from .04 to .79 with a median of .39, suggesting fairly low concensus among judges and indicting the dependability of visual inferences. However, there was no relationship found between the reliability of the judges and the degree of agreement with time-series

inferences. Jones et al. (1978) considered their findings (low agreement with high serial dependency and statistically reliable changes in level, and high agreement with low serial dependency and unreliable changes in level) to be contrary to the purpose of research using an operant methodology.

The only investigation to appear in the literature raising doubt about the validity of statistical analyses is that of Kazdin (1976). He noted that although statistical evaluation may be more reliable for a given pattern of data, it is not possible to state that it is generally more reliable. He cited an investigation by Gottman (1973) in which data were clearly significant by visual analysis, yet with the application of statistical time-series analysis, no significant effects appeared. One explanation given for this discrepancy was the short duration of individual phases, making conclusions from the time-series analysis equivocal.

It is clear from the research conducted to date that there are problems in analyzing time-series data. While traditional statistical methodology appears inappropriate, only a relatively few appropriate statistical procedures for time series data have been developed. However, the alternative, visual analysis, appears very problematic (DeProspero & Cohen, 1979; Gottman & Glass, 1978; Jones et al., 1978; Tindal, Deno, & Ysseldyke, 1983; Wampold & Furlong, 1981). Generally, this research has found visual analysis to have low reliability and to fluctuate quite dramatically with various characteristics of the data array. In an effort to improve this form of analysis, several studies have been conducted on the development of guidelines for visually

analyzing time series data. The major focus of this work has been on the use of aimlines and decision rules (Bohannon, 1975; Deno, Chiang, Tindal, & Blackburn, 1979; Liberty, 1972; Martin, 1980; Mirkin & Deno, 1979; White & Haring, 1980). Generally, for any given time series an aimline is established which begins at the current level of functioning and extends to the (goal) level of expected outcome. Subsequently, decision rules are devised for making program changes, typically when the data fall below the aimline for two or three consecutive days. Ideally, the use of such aimline/decision rules should provide systematicity to the interpretation of data, though no research has been conducted on the the effects of using aimlines/decision rules on the reliability of visual analysis.

In summary, it appears that two different strategies have been followed in attempting to develop an empirical basis for some formal type of analytic procedures: (a) investigations in the field of statistical analysis which focus on the development of appropriate and sensitive techniques for analyzing time-series data, and (b) investigations focusing on the problems of visual analysis and the development of guidelines and procedures for ameliorating these problems. The methodology of this research encompasses both strategies. Using a technique recently reported by Tryon (1982), time series data were analyzed statistically for determining treatment (intervention) effects. At the same time, the degree to which these results were in concord with visual analysis was investigated. Two evaluation components within the visual analysis of data were manipulated: the training of judges and the use of aimlines. This

study represents a reanalysis of previous research conducted by Tindal et al. (1983).

Method

Subjects

Subjects for this study were in-service and pre-service teachers from three different locations around a large midwestern city. Two of the sites were school districts, accounting for nine of the subjects, all of whom were currently teaching. Teachers in these two sites were assigned randomly to different treatment conditions, with the three subjects from one district assigned to the experimental group and the six from the other district assigned to the control group. The remaining 42 subjects were students taking a required special education class at a large midwestern university. Most of these subjects were currently teaching or former teachers. Subjects from this pool were assigned randomly to treatment groups in proportion to the number needed for bringing the experimental and control groups to approximately the same size. Twenty students were assigned to the control group and 28 were assigned to the experimental group.

Training. The training of subjects involved both an in-service workshop and a take-home training module. The teachers in the experimental group were given training in the analysis of graphed data for evaluating instructional programs. This entailed explanations and exercises in summarizing student performance and using it to make interpretations. Included in the summarization of time-series data were computations of step changes, medians, slopes (using the split-middle technique; White, 1971), variability (using total bounce;

Pennypacker, Koenig, & Lindsley, 1972), and overlap (Parsonson & Baer, 1978). The rest of the workshop was devoted to the use of this information for evaluating instruction. The teachers in the control group were given training in the development of measurement techniques in the areas of reading, writing, and spelling. They were trained in assessing students to determine performance discrepancies, sampling curriculum materials to find an appropriate instructional level, and developing a measurement system to monitor student improvement.

Both workshops lasted approximately 2½ hours. Following the workshop, the experimental materials (graphs, response sheets, and directions) for 14 graphs were distributed. Following completion of these graphs (which ranged from one week for the subjects in the class to three weeks for subjects in the schools), the second set of 14 graphs was distributed. The completion and return of this material again took one week for the subjects in the class and three weeks for those in the schools.

Experimental materials. A total of 28 different graphs were constructed in which slope and variability were systematically manipulated. Two phases were displayed in each graph--11 data points in the baseline and 15 data points in the intervention phase. A vertical line was drawn separating the two phases. The aimline represented a 30% percent improvement over the median of the last three days during baseline. To ensure comparability between the graphs with and without aimlines, the absolute level of this median value was nearly the same across both aimline conditions within each respective level of slope. Although the slope was manipulated only in

the intervention phase, variability was manipulated in both baseline and during the intervention. A total of three levels of slope and four conditions of variability were included in the graphs.

With variability manipulated in both baseline and intervention, two different combinations of variability were included: a bounce of 5 data points and one of 15 data points. For every combination of slope, variability increased (5-15), decreased (15-5), remained at the same low level (5-5) or remained at the same high level (15-15). This resulted in the following combinations of graphed data:

(a) Six graphs showed an increase in variability from baseline to intervention from 5 data points to 15 data points, with a concurrent increase in slope from 0 to 10 degrees for two graphs, an increase from 0 to 15 degrees in two graphs, and an increase from 0 to 20 degrees for the final two graphs. Of these six graphs, three had an aimline drawn in during the intervention phase, one from each combination of slope and variability.

(b) Six graphs showed a decrease in variability from 15 data points in baseline to 5 data points in the intervention phase. For two of these graphs, the change in slope from baseline to intervention involved an increase from 0 to 10 degrees, two graphs depicted an increase from 0 to 15 degrees and two had an increase from 0 to 20 degrees. Again, an aimline was drawn in on half (three) of the above graphs, one from each combination.

(c) Six graphs showed steady (unchanging) variability at a low level (5 data points bounce) from baseline to intervention. Again, the slope changed from 0 to 10 degrees on two of the graphs, 0 to 15 degrees on two graphs, and 0 to 20 degrees on the final two graphs. For each pair of slope-variability, one had an aimline and one did not.

(d) Six graphs showed steady (unchanging) variability at a high level (15 data points bounce) from baseline to intervention. Each level of slope (10, 15, and 20 degrees) was represented; two graphs displayed a change of 0 to 10 degrees, two graphs displayed a change of 0 to 15 degrees, and two showed a change of 0 to 20 degrees. Again, aimlines were present on half of these -- one in each combination of slope-variability.

The final four graphs which were constructed had the following characteristics:

(e) Four graphs that were given at time 1 were again given at time 2, with exactly the same data array depicted. All of these graphs displayed a low slope change (0 to 10 degrees) and constant variability (either the same low or same high variability). For each of the two variability conditions, one had an aimline present and one had no aimline present.

Dependent variables. As noted previously, each subject was given 14 of the graphs following training. Each graph had a response sheet containing several different questions. Responses to only the first question "Was the intervention depicted on the graph an effective one?" were attended to in this investigation. Responses to this question consisted of rating the effectiveness on a 1-4 scale, with 1 being definitely not effective and 4 being definitely effective. After the first set of 14 graphs and responses were collected, another set of 14 graphs was distributed. The order in which the graphs were organized (and completed) was determined randomly for both groups of subjects.

All subjects' responses to the original research question were recoded as dichotomous responses signifying judgments of either the presence or absence of an effective intervention. The four-point scale therefore was reduced to a two-point scale, with ones (definitely not effective) and twos (possibly effective) recoded as zeros, while threes (moderately effective) and fours (definitely effective) were recoded as ones. A zero represented the judgment of the intervention as having no effect, while a one represented a judgment of the intervention as having an effect.

All graphs were reanalyzed using the time-series analysis procedures described by Tryon (1982). This technique had been reported as particularly amenable for use on time series having a low number of data points. In this study, there were only 15 data points in the intervention phase. As Tryon (1982) noted, these statistical procedures may be used either within phases or between phases. His description for use between phases was, however, far less explicit, and indeed, the phase change was ignored in the example analysis he provided in the report. Rather than incorporating baseline data with intervention data, as done in that report, the statistical analyses conducted here involved only intervention phase data.

With both visual analyses and statistical analyses organized into dichotomous variables indicating either significant or nonsignificant intervention effects, the data were cross-tabulated using a chi-square analysis. A comparison between visual and statistical analyses was conducted for all of the major independent variables including the effect of slope, variability, training in data utilization and the use of aimlines, as well as reliability over time. No interactions were investigated in this study (see Tindal, Deno, & Ysseldyke for interaction analyses).

Results

The results of the visual and statistical analyses for all graphs are listed in Table 1. Included are the numbers of judges indicating whether the intervention was effective or not effective on the basis of visual analysis, for both trained and untrained judges, as well as both the c-statistic and the z-statistic, upon which the tests of

significance are based for time-series analysis. In the analyses which follow, the relationship between visual and statistical analyses is reported in two ways: first, whether the relationship is significant using a chi-square analysis; and second, the percentage of intervention effects that are misclassified by visual analysis, using the results of the statistical test as the criterion.

Insert Table 1 about here

The relationship between visual and statistical analyses for trained and untrained judges appears in Table 2. Although there was a significant relationship between visual and statistical analyses for trained judges, no such relationship occurred for untrained judges. Nevertheless, there was only a small difference between the two groups in the percentage of effects which were misclassified, with visual analysis of trained judges only slightly more accurate than that of untrained judges. The major difference appeared in the large number of effects that untrained judges viewed as significant, with nearly a third again as many as with trained judges. Many of these intervention effects were not statistically significant, resulting in the significant chi-square and more accurate classification for trained judges. Of the statistically significant effects, trained judges classified equal proportions as either significant or not, while untrained judges viewed far more as significant.

Insert Table 2 about here

The influence of slope of improvement on the agreement between visual and statistical analyses is summarized in Table 3. There was a significant relationship between the results of visual and statistical analyses when the slope was steep (20 degrees) or low (10 degrees), but not when it was intermediate (15 degrees). In this latter condition, the number of intervention effects that were misclassified approached nearly 50%. Although a low slope of 10 degrees resulted in a significant relationship between the two types of analyses, the percentage of misclassifications was the highest, at 56%. In contrast, the percentage misclassified was quite low (35%) when the slope was steep.

Insert Table 3 about here

The types of errors made appeared to vary as a function of the level of the slope. With a steep slope, very few judges reported nonsignificant intervention effects for effects that were statistically significant. Rather, most reported statistically nonsignificant effects as appearing to be visually significant. However, with a low slope, just the opposite occurred. More judges rated statistically significant effects as visually nonapparent. With a moderate level of slope (15 degrees), there was no difference in the types of errors made in judgment.

Little differential effect was present in the relationship between visual and statistical analyses with the use of aimlines versus no aimlines (see Table 4). Both of these conditions resulted in significant relationships between visual and statistical analyses. The most pronounced difference occurred in the rates of misclassification. Far fewer interventions were analyzed incorrectly when no aimlines were depicted on the graphs (39%) than when aimlines were present (60%). In this latter condition, not only were there more of both types of errors, but also more judgments of significance using visual analysis that were not significant statistically. The lowest frequency cell with graphs having aimlines occurred with non-significant judgments using both visual and statistical appraisals. In contrast, when graphs had no aimlines, there were approximately equal numbers of judgments of consistent decisions for both significant and nonsignificant intervention effects.

Insert Table 4 about here

Only one significant effect out of four conditions was evident for the influence of variability on the agreement between visual and statistical analyses, which occurred when variability decreased (see Table 5). When variability increased, all intervention effects were statistically nonsignificant, precluding the use of the two sample chi-square statistic. The opposite situation occurred when variability decreased, with no graphs displaying statistically non-significant effects. For each of these conditions, a one-sample chi-

square was used to analyze the degree of correspondence between visual and statistical analysis. For increasing variability, no significant relationship was found between the two types of analysis while for decreasing variability, a highly significant relationship was found. In this latter condition, nearly twice as many statistically significant interventions were deemed visually significant than those judged visually nonsignificant. This resulted in a low percentage (35%) of misclassified interventions in which statistical and visual analyses disagreed.

Insert Table 5 about here

The graphs depicting constant low variability across both pre and post intervention phases showed no significant relationship for the two analytic procedures. Finally, with constant high variability depicted in both phases, the relationship between visual and statistical analyses approached significance (p = .06). In both conditions of constant variability, as well as in the condition with increased variability, the percentage of misclassified intervention effects was around 50%.

The comparison of visual and statistical analyses in terms of the reliability from Time 1 to Time 2 was confined to slopes of 10 degrees and constant variability, either low or high (see Table 6). A large difference between the two conditions of variability was found in the significance of the relationship between visual and statistical analyses. With low variability, the relationship between the two

types of analyses was significant, while no such finding occurred for high variability. Yet, little difference existed in the percentage misclassified, with both above 50%. For both conditions, more judges rated statistically significant effects as visually nonsignificant.

Insert Table 6 about here

Discussion

The major findings of this research corroborate much of the previous research conducted in this area (Gottman & Glass, 1978; Jones et al., 1975, 1977, 1978). The agreement between visual and statistical analyses is modest at best (under ideal conditions) and otherwise quite low. In general, there was a high percentage of intervention effects that were misclassified, using the results of the statistical analysis as the criterion. That is, many interventions were viewed as significant but were not statistically significant, and vice versa (statistically significant effects were viewed as nonsignificant). The lowest percentage of misclassifications occurred under the quite ideal conditions of a relatively steep slope (of 20 degrees) or decreasing variability. Under most of the conditions, the percentage of misclassified intervention effects hovered around 50%.

Although there were several findings, specific to the manipulated variables, that were consistent with established data-utilization procedures, (Parsonson & Baer, 1978), there also were several findings in contrast with previous reports. For instance, with a steep slope there were relatively few interventions that were viewed as

nonsignificant. Only 14% of the statistically significant interventions were visually analyzed as nonsignificant. The problem appeared to be one of "seeing" effects that were not there rather than one of "not seeing" effects that were present. This finding is somewhat in contrast with the results obtained by Jones et al. (1978), in which judges were most in agreement with time series analysis when the analyses showed no effect and least in agreement when the analyses showed an effect. It is also in contrast to Baer's (1977) analysis of behavioral (n=1) research designs as possessing very low probabilities of Type I errors and correspondingly high probabilities of Type II errors. In this research, there were several conditions in which the subjects tended to visually analyze effects as significant, though statistical analysis did not corroborate the same interpretation: when the judges were untrained, when the slope was steep, when aimlines were present, and when variability was constantly high.

Agreement between visual and statistical analyses for various levels of slope revealed an interesting relationship. Only in the extreme levels (either low or high) was the relationship significant. Although visual judgment was more consistent with statistical analysis when the data array showed more pronounced change (or lack thereof), there still was a high percentage of misinterpreted effects. Apparently, extremely steep slopes or absolutely no slope, must be present before the relationship appears to be not only statistically significant, but one in which the rates of misclassification also are low.

The effect of variability on the relationship between visual and

statistical analyses generally was quite predictable. The worst case occurred with increased variability. Here, the judges were split evenly over the visual analysis of intervention effects, even though all effects had been found to be statistically nonsignificant. In contrast, when variability decreased, there was a 2 to 1 margin of judges viewing the effects as visually significant, which was in agreement with the statistical analysis. The effect of low constant variability resulted in a nearly even split, similar to that found with increasing variability. While most of the interventions depicted in graphs with constant high variability were statistically nonsignificant, the majority of judges viewed the effects as significant. This finding is similar to the previous finding reported by Tindal et al. (1983), in which some degree of variability is viewed as indicative of treatment effects.

The fact that a higher (more significant) relationship between visual and statistical analyses was obtained for trained judges than untrained judges is consistent with the predicted treatment effect. Apparently, some training effect, albeit modest, occurred through the use of a two hour lecture and a take-home manual. Previous doubts concerning the efficacy of training had been expressed by Tindal et al. (1983). This doubt primarily was a result of the low reliability obtained by both trained and untrained judges. Although judgments of trained judges showed a higher relationship with statistical analysis, it is also true that the levels of misclassification were quite high and only slightly improved over that obtained by untrained judges (46% vs 50%).

The most interesting difference between the trained and untrained judges was in the types of errors made. Generally, untrained judges tended to classify more interventions as significant. This resulted in a higher percentage of accurately classified significant effects by nearly 10% (50% for trained and 60% for untrained judges). However, it appears that the judgments of significant effects made by untrained judges were higher in general. Again, this is consistent with the previous research by Tindal et al. (1983). The implication of this finding is that a higher percentage of Type I errors are made by untrained judges - judgments of an effect being made when in fact no effect is present - using statistical analysis as the criterion. Furthermore, the inherently low probability of Type I error in visual analysis cannot be assumed, but rather occurs as a function of training.

In contrast to the findings for trained vs untrained judges, the use of aimlines appeared not only to be ineffective, but actually to interfere with judgments of effects. With aimlines present, more of the statistically nonsignificant interventions were deemed effective. That is, the aimline appeared to sway judgments of effects when none were actually present. Although aimlines may simplify the decision-making process, it is also possible that their use may distort that judgment. The problem with the aimline/decision rule system may reside in the fact that critical data (i.e., slope and variability) essentially are ignored. This problem may be remediated by the use of an aimline/decision rule similar to that developed by Mirkin, Deno, Fuchs, Wesson, Tindal, Marston, and Kuehnle (1981), in which program

analysis is based on the degree to which the slope of student improvement intersects with the aimline of the program.

The analysis of reliability of agreement appeared as predicted: With low constant variability a significant relationship occurred, while no such relationship occurred for high constant variability. However, the percentage of misclassified effects was quite high for both groups. Again, more judgments of significant effects using visual analysis were reported when the variability was high.

In summary, it appears that visual analysis of time series data shows very modest agreement with statistical analysis, and is influenced in quite predictable ways by characteristics of the data. The use of aimlines does not adequately solve the problems of visual analysis and may well exacerbate them. Although training appears necessary for establishing the statistical conclusion validity (Cook & Campbell, 1979) of visual analysis, some type of decision rule system also must be developed, given the high rates of misclassification of effects, using statistical analysis as the criterion.

References

Baer, D. M. Perhaps it would be better not to know everything. Journal of Applied Behavior Analysis, 1977, 10, 167-172.

Bohannon, R. Direct and daily measurement procedures in the identification and treatment of reading behaviors of children in special education. Unpublished Doctoral Dissertation, University of Washington, 1975.

Cook, T. D., & Campbell, D. T. Quasi-experimentation: Design and analysis issues for field settings. Chicago: Rand McNally, 1979.

Deno, S., Chiang, B., Tindal, G., & Blackburn, M. Experimental analysis of program components: An approach to research in CSDCs (Research Report No. 12). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1979.

DeProspero, A., & Cohen, S. Inconsistent visual analysis of intrasubject data. Journal of Applied Behavior Analysis, 1979, 12, 573-579.

Edgington, E. S. Statistical inference from N=1 experiments. The Journal of Psychology, 1967, 65, 195-199.

Edgington, E. S. N=1 experiments: Hypothesis testing. The Canadian Psychologist, 1972, 2, 121-13[illegible] (a)

Edgington, E. S. The design of one subject experiments for testing hypotheses. Western Psychologist, 1972, 3, 33-38. (b)

Fisher, R. A. The design of experiments (6th ed). Edinburgh: Oliver & Boyd, 1951.

Glass, G. V., Willson, V. L., & Gottman, J. M. Design and analysis of time-series experiments. Boulder, Colo.: University of Colorado Press, 1975.

Gottman, J. M. N-of-one and N-of-two research in psycho-therapy. Psychological Bulletin, 1973, 80, 93-105.

Gottman, J. M., & Glass, G. V. Analysis of interrupted time-series experiments. In T. R. Kratochwill (Ed.), Single subject research: Strategies for evaluating change. New York: Academic Press, 1978.

Hall, R. V., Fox, R., Willard, D., Goldsmith, L., Emerson, M., Owen, M., Davis, F., & Porcia, E. The teacher as observer and experimenter in the modification of disputing and talking out behaviors. Journal of Applied Behavior Analysis, 1971, 4, 141-149.

Hartmann, D. P. Forcing square pegs into holes: Some comments on "An analysis of variance model for intra-subject replication design." Journal of Applied Behavior Analysis, 1974, 7, 635-638.

Jones, R. R., Vaught, R. S., & Weinrott, M. Time series analysis in operant research. Journal of Applied Behavior Analysis, 1977, 10, 151-166.

Jones, R. R., Weinrott, M., & Vaught, R. S. Visual vs. statistical inference in operant research. In A. E. Kazdin (Ed.), The use of statistics in N=1 research. Symposium presented at the annual convention of the American Psychological Association, Chicago, September, 1975.

Jones, R. R., Weinrott, M. R., & Vaught, R. S. Effects of serial dependency on the agreement between visual and statistical inference. Journal of Applied Behavior Analysis, 1978, 11, 272-283.

Kazdin, A. E. Behavior modification in applied settings. Horewood, Ill.: Dorsey Press, 1975.

Kazdin, A. E. Statistical analyses of single-case experimental designs. In M. Hersen & D. Barlow (Eds.), Single case experimental designs: Strategies for studying behavior change. New York: Pergamon Press, 1976.

Kratochwill, T. R., Alden, K., Demuth, D., Dawson, D., Panicucci, C., Arntson, P., McMurray, N., Hempstead, J., & Levin, J. A further consideration in the application of an analysis-of-variance model for the intra-subject replication design. Journal of Applied Behavior Analysis, 1974, 7, 629-633.

Liberty, K. A. Data decision rules. Unpublished working paper #20. Eugene, Ore.: Regional Resources Center, University of Oregon, 1972.

Martin, M. A. A comparison of variations in data utilization procedures on the reading performance of mildly handicapped students. Unpublished PhD dissertation. University of Washington, 1980.

Mirkin, P. K., & Deno, S. L. Formative evaluation in the classroom: An approach to improving instruction (Research Report No. 10). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1979.

Mirkin, P. K., Deno, S., Fuchs, L., Wesson, C., Tindal, G., Marston, D., & Kuehnle, K. Procedures to develop and monitor progress on IEP goals. Minneapolis: Institute for Research on Learning Disabilities, 1981.

Parsonson, B. S., & Baer, D. M. The analysis and presentation of graphic data. In T. R. Kratochwill (Ed.), Single subject research: Strategies for evaluating change. New York: Academic Press, 1978.

Pennypacker, H. S., Koenig, C. H., & Lindsley, O. R. Handbook of the standard behavior chart (prelim. ed.). Kansas City, KS: Precision Media, 1972.

Revusky, S. H. Some statistical treatments compatible with individual organism methodology. Journal of Experimental Analysis of Behavior, 1967, 10, 319-330.

Scheffe, H. The analysis of variance. New York: Wiley, 1959.

Tindal, G., Deno, S., & Ysseldyke, J. Visual analysis of time series data: Factors of influence and level of reliability (Research Report No. 112). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1983.

Tryon, W. W. A simplified time-series analysis for evaluating treatment interventions. Journal of Applied Behavior Analysis, 1982, 15(3), 423-429.

Wampold, B. E., & Furlong, M. J. The heuristics of visual inference. Behavioral Assessment, 1981, 3, 79-92.

White, O. R. A pragmatic approach to the description of progress in the single case. University of Oregon: Unpublished Doctoral Dissertation, 1971.

White, O. R., & Haring, N. G. Exceptional teaching (2nd ed.). Columbus, Ohio: Charles E. Merrill, 1980.

Table 1

Results of Visual and Statistical Analyses for All Graphs

Graph Number	Trained Judges		Untrained Judges		Statistic	
	Effective	Not Effective	Effective	Not Effective	c	z
1*	21	4	24	2	.80	3.31
2	11	14	15	11	.28	1.14
3*	15	10	22	4	.40	1.66
4*	6	19	13	13	.41	1.87
5*	25	0	26	0	.86	3.58
6	5	20	5	21	.17	0.71
7	14	11	17	9	.09	0.38
8	6	19	12	14	-.76	-0.32
9	22	3	24	2	.17	0.69
10*	4	21	13	13	.82	3.40
11*	5	20	2	24	.53	2.19
12*	25	0	26	0	.81	3.35
13	22	3	26	0	.29	1.21
14*	1	24	5	21	.55	2.27
15*	24	1	21	5	.53	2.21
16	7	18	20	6	.28	1.16
17*	14	11	15	11	.45	1.87
18	4	21	7	19	.22	0.92
19*	2	23	6	20	.55	2.27
20	22	3	23	3	.22	0.91
21*	8	17	18	8	.72	2.89
22	15	10	26	0	.11	0.47
23	6	19	11	15	.13	0.56
24	1	24	9	17	.26	1.08
25	6	19	8	18	.27	1.13
26*	8	17	19	7	.66	2.75
27*	17	8	17	9	.64	2.66
28	7	18	14	12	.35	1.40

*Statistically significant at $p \leq .05$.

Table 2

Relationship Between Results of Visual and Statistical Analyses for Trained and Untrained Judges

Judges	Visual Analysis Results	Statistical Analysis Results: Significant	Statistical Analysis Results: Nonsignificant
Trained[a]			
	Significant	188	135
	Nonsignificant	187	190
Untrained[b]			
	Significant	233	211
	Nonsignificant	157	127

[a] $n = 700$; $\chi^2 = 4.84$, $p \leq .05$; % misclassified = 46.

[b] $n = 728$; $\chi^2 = .43$, $p \leq .50$; % misclassified = 53.

Table 3

Relationship Between Results of Visual and Statistical Analyses for Graphs with Slopes of 10^{o}, 15^{o}, and 20^{o}

Slope	Visual Analysis Results	Statistical Analysis Results: Significant	Statistical Analysis Results: Nonsignificant
10^{o}[a]			
	Significant	77	101
	Nonsignificant	127	103
15^{o}[b]			
	Significant	106	92
	Nonsignificant	98	112
20^{o}[c]			
	Significant	219	105
	Nonsignificant	36	48

[a] $n = 408$; $\chi^2 = 5.27$, $p \leq .03$; % misclassified = 56.

[b] $n = 408$; $\chi^2 = 1.66$, $p \leq .20$; % misclassified = 47.

[c] $n = 408$; $\chi^2 = 16.37$, $p \leq .001$; % misclassified = 35.

Table 4

Relationship Between Results of Visual and Statistical Analyses for Graphs with and without Aimlines

Condition	Visual Analysis Results	Statistical Analysis Results: Significant	Statistical Analysis Results: Nonsignificant
With Aimlines[a]			
	Significant	146	207
	Nonsignificant	160	99
Without Aimlines[b]			
	Significant	195	125
	Nonsignificant	111	181

[a] $n = 612$; $\chi^2 = 24.10$, $p \leq .001$; % misclassified = 60.

[b] $n = 612$ $\chi^2 = 31.18$, $p \leq .001$; % misclassified = 39.

Table 5

Relationship Between Results of Visual and Statistical Analyses for Four Variability Conditions

Variability Condition	Visual Analysis Results	Statistical Analysis Results: Significant	Statistical Analysis Results: Nonsignificant
Increasing[a]			
	Significant	--	152
	Nonsignificant	--	152
Decreasing[b]			
	Significant	198	--
	Nonsignificant	106	--
Constant Low[c]			
	Significant	129	19
	Nonsignificant	126	32
Constant High[d]			
	Significant	37	145
	Nonsignificant	14	110

[a] $n = 304$; $\chi^2 = 0$, $p \leq 1.0$; % misclassified = 50.

[b] $n = 304$; $\chi^2 = 27.87$, $p \leq .001$; % misclassified = 35.

[c] $n = 306$; $\chi^2 = 2.52$, $p \leq .13$; % misclassified = 47.

[d] $n = 306$; $\chi^2 = 3.71$, $p \leq .06$; % misclassified = 52.

Table 6

Relationship Between Results of Visual and Statistical Analyses from Time 1 to Time 2 for Graphs with Low Slopes (10^{o}) and Low or High Constant Variability

Variability	Visual Analysis Results	Statistical Analysis Results: Significant	Statistical Analysis Results: Nonsignificant
Low Constant[a]			
	Significant	25	45
	Nonsignificant	77	57
High Constant[b]			
	Significant	37	48
	Nonsignificant	65	54

[a] $n = 204$; $\chi^2 = 7.85$, $p \leq .01$; % misclassified = 60.

[b] $n = 204$; $\chi^2 = 2.02$, $p \leq .16$; % misclassified = 55.

PUBLICATIONS

Institute for Research on Learning Disabilities
University of Minnesota

The Institute is not funded for the distribution of its publications. Publications may be obtained for $4.00 each, a fee designed to cover printing and postage costs. Only checks and money orders payable to the University of Minnesota can be accepted. All orders must be prepaid. Requests should be directed to: Editor, IRLD, 350 Elliott Hall; 75 East River Road, University of Minnesota, Minneapolis, MN 55455.

The publications listed here are only those that have been prepared since 1982. For a complete, annotated list of all IRLD publications, write to the Editor.

Wesson, C., Mirkin, P., & Deno, S. Teachers' use of self instructional materials for learning procedures for developing and monitoring progress on IEP goals (Research Report No. 63). January, 1982.

Fuchs, L., Wesson, C., Tindal, G., Mirkin, P., & Deno, S. Instructional changes, student performance, and teacher preferences: The effects of specific measurement and evaluation procedures (Research Report No. 64). January, 1982.

Potter, M., & Mirkin, P. Instructional planning and implementation practices of elementary and secondary resource room teachers: Is there a difference? (Research Report No. 65). January, 1982.

Thurlow, M. L., & Ysseldyke, J. E. Teachers' beliefs about LD students (Research Report No. 66). January, 1982.

Graden, J., Thurlow, M. L., & Ysseldyke, J. E. Academic engaged time and its relationship to learning: A review of the literature (Monograph No. 17). January, 1982.

King, R., Wesson, C., & Deno, S. Direct and frequent measurement of student performance: Does it take too much time? (Research Report No. 67). February, 1982.

Greener, J. W., & Thurlow, M. L. Teacher opinions about professional education training programs (Research Report No. 68). March, 1982.

Algozzine, B., & Ysseldyke, J. Learning disabilities as a subset of school failure: The oversophistication of a concept (Research Report No. 69). March, 1982.

Fuchs, D., Zern, D. S., & Fuchs, L. S. A microanalysis of participant behavior in familiar and unfamiliar test conditions (Research Report No. 70). March, 1982.

Shinn, M. R., Ysseldyke, J., Deno, S., & Tindal, G. A comparison of psychometric and functional differences between students labeled learning disabled and low achieving (Research Report No. 71). March, 1982.

Thurlow, M. L. Graden, J., Greener, J. W., & Ysseldyke, J. E. Academic responding time for LD and non-LD students (Research Report No. 72). April, 1982.

Graden, J., Thurlow, M., & Ysseldyke, J. Instructional ecology and academic responding time for students at three levels of teacher-perceived behavioral competence (Research Report No. 73). April, 1982.

Algozzine, B., Ysseldyke, J., & Christenson, S. The influence of teachers' tolerances for specific kinds of behaviors on their ratings of a third grade student (Research Report No. 74). April, 1982.

Wesson, C., Deno, S., & Mirkin, P. Research on developing and monitoring progress on IEP goals: Current findings and implications for practice (Monograph No. 18). April, 1982.

Mirkin, P., Marston, D., & Deno, S. L. Direct and repeated measurement of academic skills: An alternative to traditional screening, referral, and identification of learning disabled students (Research Report No. 75). May, 1982.

Algozzine, B., Ysseldyke, J., Christenson, S., & Thurlow, M. Teachers' intervention choices for children exhibiting different behaviors in school (Research Report No. 76). June, 1982.

Tucker, J., Stevens, L. J., & Ysseldyke, J. E. Learning disabilities: The experts speak out (Research Report No. 77). June, 1982.

Thurlow, M. L., Ysseldyke, J. E., Graden, J., Greener, J. W., & Mecklenberg, C. Academic responding time for LD students receiving different levels of special education services (Research Report No. 78). June, 1982.

Graden, J. L., Thurlow, M. L., Ysseldyke, J. E., & Algozzine, B. Instructional ecology and academic responding time for students in different reading groups (Research Report No. 79). July, 1982.

Mirkin, P. K., & Potter, M. L. A survey of program planning and implementation practices of LD teachers (Research Report No. 80). July, 1982.

Fuchs, L. S., Fuchs, D., & Warren, L. M. Special education practice in evaluating student progress toward goals (Research Report No. 81). July, 1982.

Kuehnle, K., Deno, S. L., & Mirkin, P. K. Behavioral measurement of social adjustment: What behaviors? What setting? (Research Report No. 82). July, 1982.

Fuchs, D., Dailey, Ann Madsen, & Fuchs, L. S. Examiner familiarity and the relation between qualitative and quantitative indices of expressive language (Research Report No. 83). July, 1982.

Videen, J., Deno, S., & Marston, D. Correct word sequences: A valid indicator of proficiency in written expression (Research Report No. 84). July, 1982.

Potter, M. L. Application of a decision theory model to eligibility and classification decisions in special education (Research Report No. 85). July, 1982.

Greener, J. E., Thurlow, M. L., Graden, J. L., & Ysseldyke, J. E. The educational environment and students' responding times as a function of students' teacher-perceived academic competence (Research Report No. 86). August, 1982.

Deno, S., Marston, D., Mirkin, P., Lowry, L., Sindelar, P., & Jenkins, J. The use of standard tasks to measure achievement in reading, spelling, and written expression: A normative and developmental study (Research Report No. 87). August, 1982.

Skiba, R., Wesson, C., & Deno, S. L. The effects of training teachers in the use of formative evaluation in reading: An experimental-control comparison (Research Report No. 88). September, 1982.

Marston, D., Tindal, G., & Deno, S. L. Eligibility for learning disability services: A direct and repeated measurement approach (Research Report No. 89). September, 1982.

Thurlow, M. L., Ysseldyke, J. E., & Graden, J. L. LD students' active academic responding in regular and resource classrooms (Research Report No. 90). September, 1982.

Ysseldyke, J. E., Christenson, S., Pianta, R., Thurlow, M. L., & Algozzine, B. An analysis of current practice in referring students for psychoeducational evaluation: Implications for change (Research Report No. 91). October, 1982.

Ysseldyke, J. E., Algozzine, B., & Epps, S. A logical and empirical analysis of current practices in classifying students as handicapped (Research Report No. 92). October, 1982.

Tindal, G., Marston, D., Deno, S. L., & Germann, G. Curriculum differences in direct repeated measures of reading (Research Report No. 93). October, 1982.

Fuchs, L.S., Deno, S. L., & Marston, D. Use of aggregation to improve the reliability of simple direct measures of academic performance (Research Report No. 94). October, 1982.

Ysseldyke, J. E., Thurlow, M. L., Mecklenburg, C., & Graden, J. Observed changes in instruction and student responding as a function of referral and special education placement (Research Report No. 95). October, 1982.

Fuchs, L. S., Deno, S. L., & Mirkin, P. K. Effects of frequent curriculum-based measurement and evaluation on student achievement and knowledge of performance: An experimental study (Research Report No. 96). November, 1982.

Fuchs, L. S., Deno, S. L., & Mirkin, P. K. Direct and frequent measurement and evaluation: Effects on instruction and estimates of student progress (Research Report No. 97). November, 1982.

Tindal, G., Wesson, C., Germann, G., Deno, S. L., & Mirkin, P. K. The Pine County model for special education delivery: A data-based system (Monograph No. 19). November, 1982.

Epps, S., Ysseldyke, J. E., & Algozzine, B. An analysis of the conceptual framework underlying definitions of learning disabilities (Research Report No. 98). November, 1982.

Epps, S., Ysseldyke, J. E., & Algozzine, B. Public-policy implications of different definitions of learning disabilities (Research Report No. 99). November, 1982.

Ysseldyke, J. E., Thurlow, M. L., Graden, J. L., Wesson, C., Deno, S. L., & Algozzine, B. Generalizations from five years of research on assessment and decision making (Research Report No. 100). November, 1982.

Marston, D., & Deno, S. L. Measuring academic progress of students with learning difficulties: A comparison of the semi-logarithmic chart and equal interval graph paper (Research Report No. 101). November, 1982.

Beattie, S., Grise, P., & Algozzine, B. Effects of test modifications on minimum competency test performance of third grade learning disabled students (Research Report No. 102). December, 1982

Algozzine, B., Ysseldyke, J. E., & Christenson, S. An analysis of the incidence of special class placement: The masses are burgeoning (Research Report No. 103). December, 1982.

Marston, D., Tindal, G., & Deno, S. L. Predictive efficiency of direct, repeated measurement: An analysis of cost and accuracy in classification (Research Report No. 104). December, 1982.

Wesson, C., Deno, S., Mirkin, P., Sevcik, B., Skiba, R., King, R., Tindal, G., & Maruyama, G. Teaching structure and student achievement effects of curriculum-based measurement: A causal (structural) analysis (Research Report No. 105). December, 1982.

Mirkin, P. K., Fuchs, L. S., & Deno, S. L. (Eds.). Considerations for designing a continuous evaluation system: An integrative review (Monograph No. 20). December, 1982.

Marston, D., & Deno, S. L. Implementation of direct and repeated measurement in the school setting (Research Report No. 106). December, 1982.

Deno, S. L., King, R., Skiba, R., Sevcik, B., & Wesson, C. The structure of instruction rating scale (SIRS): Development and technical characteristics (Research Report No. 107). January, 1983.

Thurlow, M. L., Ysseldyke, J. E., & Casey, A. Criteria for identifying LD students: Definitional problems exemplified (Research Report No. 108). January, 1983.

Tindal, G., Marston, D., & Deno, S. L. The reliability of direct and repeated measurement (Research Report No. 109). February, 1983.

Fuchs, D., Fuchs, L. S., Dailey, A. M., & Power, M. H. Effects of pre-test contact with experienced and inexperienced examiners on handicapped children's performance (Research Report No. 110). February, 1983

King, R. P., Deno, S., Mirkin, P., & Wesson, C. The effects of training teachers in the use of formative evaluation in reading: An experimental-control comparison (Research Report No. 111). February, 1983.

Tindal, G., Deno, S. L., & Ysseldyke, J. E. Visual analysis of time series data: Factors of influence and level of reliability (Research Report No. 112). March, 1983.

Tindal, G, Shinn, M., Fuchs, L., Fuchs, D., Deno, S., & Germann, G. The technical adequacy of a basal reading series mastery test (Research Report No. 113). April, 1983.

Sevcik, B., Skiba, R., Tindal, G., King, R., Wesson, C., Mirkin, P., & Deno, S. Communication of IEP goals and student progress among parents, regular classroom teachers, and administrators using systematic formative evaluation (Research Report No. 114). April, 1983.

Wesson, C. Two student self-management techniques applied to data-based program modification (Research Report No. 115). April, 1983.

Wesson, C., Skiba, R., Sevcik, B., King, R., Tindal, G., Mirkin, P., & Deno, S. The impact of the structure of instruction and the use of technically adequate instructional data on reading improvement (Research Report No. 116). May, 1983.

Wesson, C. Teacher vs student selection of instructional activities (Research Report No. 117). May, 1983.

Tindal, G., & Deno, S. Factors influencing the agreement between visual and statistical analyses of time series data (Research Report No. 118). June, 1983.